BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Problem Image Mailbox.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N

A2

(11) International Publication Number: WO 97/20031

(43) International Publication Date: 5 June 1997 (05.06.97)

(21) International Application Number: PCT/US96/19038

(22) International Filing Date: 29 November 1996 (29.11.96)

(30) Priority Data:
566,190 1 December 1995 (01.12.95) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/566,190 (CIP)
Filed on 1 December 1995 (01.12.95)

(71) Applicant *(for all designated States except US)*: YALE UNIVERSITY [US/US]; 451 College Street, New Haven, CT 06511 (US).

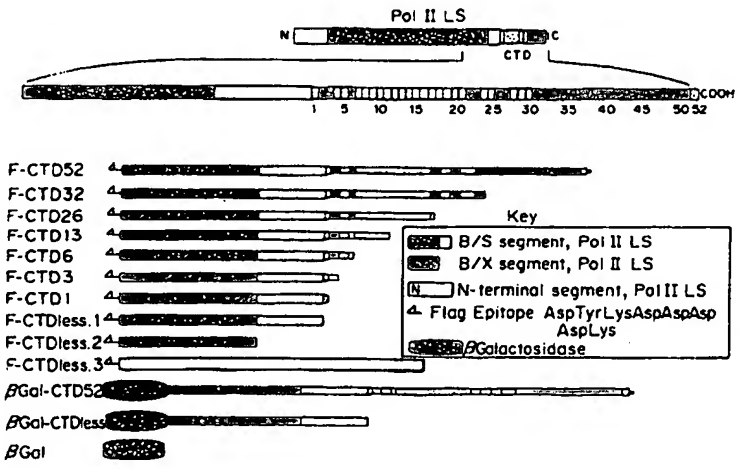
(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: WARREN, Stephen, L. [US/US]; 514 Treat Lane, Orange, CT 06477 (US).

(74) Agent: PABST, Patrea, L.; Arnall Golden & Gregory L.L.P., 2800 One Atlantic Center, 1201 West Peachtree Street, Atlanta, GA 30309-3450 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: RNA POLYMERASE II CARBOXYL TERMINAL DOMAIN-DERIVED PEPTIDES



(57) Abstract

The targeting molecules described herein are derived from the RNA polymerase II carboxyl terminus, consisting of at least two to three hexapeptide repeats of the structure: $YX_1PX_2X_3PX_4$, where Y is tyrosine, P is proline, X can be any amino acid, and X_1, X_2 and X_3 are most preferably serine or threonine, covalently conjugated to a bioactive molecule, most preferably an oligonucleotide, preferably via a linker consisting of one to two amino acids or a carbon chain of equivalent length attached using amide or disulfide chemistry to the tyrosine at the N-terminus of the peptide, leaving a free carboxyl. The peptide can be phosphorylated to alter the association with certain molecules in the nucleus, such as proteins having a serine/arginine motif and Sm snRNPs. For example, it is demonstrated that phosphorylated CTD-derived peptides bind to these nuclear proteins associated with transcription and splicing.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

RNA POLYMERASE II CARBOXYL TERMINAL DOMAIN-DERIVED PEPTIDES

Field of the Invention

This relates to the fields of immunology and protein biochemistry and more particularly relates to RNA polymerase II carboxy terminus domain-derived peptides useful for nuclear targeting of bioactive compounds, especially oligonucleotides.

Statement Regarding Federally Sponsored Research

The United States government has rights in this invention by virtue of National Institutes of Health Grant No. K08-CA01339 to Stephen L. Warren.

Background of the Invention

The notion that genes might be replaced or specifically inhibited seems realistic given the recent explosion in human genome research; the spectacular successes in transgenic animal technology; the development of viral and non-viral *ex vivo* gene transfer techniques; and the intermittent successes of antisense oligodeoxynucleotide and ribozyme technologies. The gene therapy concept has led to the formation of many biotech companies specializing in gene transfer, antisense oligonucleotides and catalytic RNAs. The pharmaceutical industry and the NIH have also made significant financial commitments to develop these technologies. Millions of dollars of investment capital and grant funds have been spent on gene therapy research, and several human gene therapy clinical trials are underway. Despite all of this excitement, genetic therapy is still in a very early stage of development. It is clear that some very difficult engineering problems must be overcome before gene transfer, antisense and catalytic RNA technologies will be used to treat human diseases.

Technical barriers can be divided into three categories: Reagent design: optimization of nucleic acid activity through gene transfer and antisense oligodeoxynucleotides and catalytic DNAs; Delivery: formulating and delivering nucleic acids into cells; and Targeting: ensuring that nucleic acid reagents are bioavailable after they enter the cell. Some of the issues that must be addressed in targeting and delivery include: a substantial fraction of the reagent (transgene/oligo/catalytic RNA) must enter the nucleus; intranuclear targeting of nucleic acid reagents must be optimized and intranuclear sequestration must be avoided; if stable expression is desired, transgenes must recombine with chromosomes; stable and transient transgenes must be accessible to the transcription machinery; and oligos and catalytic RNAs must gain access to their pre-mRNA targets.

Gene therapy research has emphasized reagent design and delivery, but not targeting. To tackle the problems of reagent design and delivery, the gene therapy researchers have drawn from a vast body of research on gene regulation, nucleic acid biochemistry, virology and membrane biology. Significant advances have been made in the understanding of genomic organization and chromatin structure; RNA polymerase II-mediated transcription; packaging and splicing of pre-mRNA; stability and translational efficiency of mRNA; kinetics and thermodynamics of nucleic acid hybridization; catalytic RNAs; viral 'vectorology' and liposome-mediated transfer of nucleic acids across cell membranes. Based upon this strong fund of knowledge, some of the problems associated with reagent design and delivery have been successfully addressed. The problem of targeting--i.e. concentrating the reagent in the appropriate

subnuclear compartment--has received far less attention. One reason may be that reagent design and delivery are perceived as more tractable problems. It seems relatively straightforward to increase the Tm of an oligonucleotide; to optimize the composition of a liposome; or to modify the enhancer in a plasmid, changes which can be rapidly assessed *in vitro*. On the other hand, it seems difficult to manipulate or even to investigate the fate of plasmids, antisense oligos or catalytic RNAs after they cross the plasma cell membrane.

As described in PCT/US95/15683 by Yale University, a macromolecular delivery method that utilizes a series of peptides with unique and versatile nuclear targeting properties has been developed, where the peptides are derived from the COOH terminal domain (CTD) of the largest subunit of RNA polymerase II and include heptapeptide units similar or identical to the following consensus sequence: Tyrosine--Serine--Proline--Threonine--Serine--Proline--Serine $(YSPTSPS)_x$. When expressed *in vivo*, the CTD peptides are phosphorylated and they accumulate in discrete compartments within the nucleus. The CTD peptides concentrate indicator molecules in discrete subnuclear compartments where pre-mRNA molecules are synthesized and spliced. The length and composition of the CTD peptides can be manipulated to obtain different intranuclear partitioning properties. The CTD peptides are functional in the nuclei of *S. cerevisiae*, *S. pombe*, nematodes, insects, plants, and all vertebrates. Since the CTD peptides accumulate precisely in discrete sites inhabited by RNA polymerase II and the spliceosomes, they should be useful in genetic therapy technologies. CTD peptides can concentrate antisense oligonucleotides, catalytic RNAs and transgenes in

the nuclear compartment where the pre-mRNAs are synthesized and processed. The CTD peptides should minimize intranuclear sequestration of therapeutic polynucleotides.

Since it is desirable to optimize the efficacy of these peptides, as well as to produce a cost-effective product, further studies to determine how many hexapeptide repeats are required for targeting, and how much variability within the repeats, are essential.

It is therefore an object of the present invention to provide a therapeutic tool for the delivery of therapeutic agents from the surface of a cell to the cell nucleus or from a cell receptor to a targeted gene, and in particular to target therapeutics to specific intranuclear regions of the cells where they are bioavailable using RNA polymerase II carboxyl terminus domain-derived (CTD-derived) peptides.

Summary of the Invention

The targeting molecules described herein are derived from the RNA polymerase II carboxyl terminus, consisting of at least two to three hexapeptide repeats of the structure: $YX_1PX_2X_3PX_4$, where Y is tyrosine, P is proline, X can be any amino acid, and X_1, X_2 and X_3 are most preferably serine or threonine, covalently conjugated to a bioactive molecule, most preferably an oligonucleotide, preferably via a linker consisting of one to two amino acids or a carbon chain of equivalent length attached using amide or disulfide chemistry to the tyrosine at the N-terminus of the peptide, leaving a free carboxyl.

The peptide can be phosphorylated to alter the association with certain molecules in the nucleus, such as proteins having a serine/arginine motif and

Sm snRNPs. For example, it is demonstrated that phosphorylated CTD-derived peptides bind to these nuclear proteins associated with transcription and splicing. Accordingly, for delivery of molecules which are desired to be in close association with RNA, it may be desirable to phosphorylate the peptide, and for delivery of molecules which are not desired to be in close association with RNA, it may be desirable to leave the peptide unphosphorylated.

Brief Description of the Drawings

Figure 1 is a schematic of fusion proteins derived from Pol II's CTD. The largest subunit of RNA Polymerase II (Pol II LS) is illustrated schematically (top). An expanded view of the CTD shows 52 heptapeptide repeats represented by variably shaded boxes. Lightly shaded boxes represent consensus repeats (YSPTSPS), and more darkly shaded boxes represent variant repeats. The CTD coding sequence was unidirectionally truncated from the C-terminus and recombinantly fused to the Flag® peptide (Flag symbol) or βGalactosidase (oval symbol). The resulting fusion proteins are described by nomenclature that begins with the N-terminus and ends with the C-terminus, including the number of heptapeptide repeats. Key: partially shaded box, B/S segment, Pol II LS; shaded box, B/X segment, Pol II LS; box containing N, N-terminal segment, Pol II LS; ≤ flag epitope AspTyrLysAspAspAspAspLys; and shaded oval, β-galactosidase.

Figure 2 is a graph of the relationship between CTD length and disruptive effect on B1C8 speckles.

CV1 cells were transfected with plasmids encoding the fusion proteins listed above the histogram

bars. Two days later, the cells were fixed and double stained with antibodies directed at the Flag^R epitope and B1C8. The pattern of B1C8 staining in each transfected cell nucleus was scored as "intact" (20-50 prominent speckles) or "disrupted" (diffuse pattern or diminutive speckles). Data were pooled from multiple experiments performed on different days. 150-250 nuclei were scored for each plasmid.

Figures 3A and 3B are schematics of plasmids expressing human β-globin transcripts and CTD-derived fusion proteins. Figure 3A is a schematic of a wild type human β-globin gene with a downstream SV40 enhancer (SV40E) inserted into an EcoRV site in multiple plasmids that express Flag-tagged proteins or βGal (Figure 1). For brevity the illustration depicts the insertion of various protein-encoding sequences into a site upstream of the β-globin gene. The Flag-tagged proteins and βGal coding sequences are under the control of the CMV promoter (CMVp). β-globin introns are represented by open boxes, exons by black boxes and noncoding flanking sequences open boxes at the ends of the gene. β-globin and CMV promoters are indicated by bent arrows. The resulting constructs are generically termed "FusionProteinβ-globin [+]." The plus sign indicates that the two genes are oriented in the same direction. The primers (P1 and P2) hybridize with complementary (cDNA) sequences within exons 1 and 2, respectively. PCR amplification with P1 and P2 yields 170 nt and 300 nt DNA fragments corresponding to spliced and unspliced transcripts, respectively. The 343 nt RNA probe used for RNase protection is shown below the β-globin gene. The open box on this probe represents non-hybridizing portion derived from pBluescript, and the black bar hybridizes with a

276 nt segment of the unspliced β-globin transcript. The 276 nt segment spans an intron-exon boundary including 203 nucleotides of exon 2 and 73 nucleotides of intron 1. Therefore, the spliced and unspliced β-globin transcripts protect 203 and 276 nucleotide segments of the probe, respectively.

Figure 3B is a schematic of a wild type human β-globin gene with a downstream SV40 enhancer (SV40E) inserted in the opposite orientation at the EcoRV site in the plasmids expressing Flag-tagged proteins or βGal. The resulting constructs are generically termed "FusionProteinβ-globin [-]." The minus sign indicates that the two genes are oriented in the opposite direction. For convenience, the protein-encoding sequences are not shown. Figure 3C is a schematic of a thalassemic human β-globin gene with a downstream SV40 enhancer (SV40E) inserted in the positive orientation into the EcoRV site in the plasmids expressing Flag-tagged proteins or βGal. The resulting constructs are generically termed "FusionProteinβ-globin^{thal} [+]." The thalassemic allele is mutated at first residue of intron 1 (G to A transition) (delta symbol). Splicing of exons 1 and 2 is achieved by utilizing three cryptic 5' splice sites and the normal 3' splice site. The oligonucleotide used for RNase protection spans the 3' splice site, but it is downstream of the cryptic 5' splice sites. Therefore, all three variably spliced transcripts register as 203 nucleotide RNAs in the RNase protection assay. For convenience, the protein-encoding sequences are not shown.

Detailed Description of the Invention

During the course of investigating the

function of the carboxy terminal doamin (CTD) of human RNA polymerase II (Pol II), including the 52 repeats of the consensus heptapeptide, tyrosine-serine-proline-threonine-serine-proline-serine (YSPTSPS), evidence has been obtained that one function of the highly repetitive structure is to link the processes Pol II transcription and pre-mRNA splicing.

Sm snRNPs and SerArg (SR) family proteins co-immunoprecipitate with Pol II molecules containing a hyperphosphorylated CTD. The association between Pol IIo and splicing factors is maintained in the absence of pre-mRNA, and the polymerase need not be transcriptionally engaged. The latter findings led to the hypothesis that a phosphorylated form of the CTD interacts with pre-mRNA splicing components *in vivo*. To test this idea, a nested set of CTD-derived proteins was assayed for the ability to alter the nuclear distribution of splicing factors, and to interfere with splicing *in vivo*. Proteins containing heptapeptides 1-52 (CTD52), 1-32 (CTD32), 1-26 (CTD26), 1-13 (CTD13), 1-6 (CTD6), 1-3 (CTD3) or 1 (CTD1) were expressed in mammalian cells. The CTD-derived proteins become phosphorylated *in vivo*, and accumulate in the nucleus even though they lack a conventional nuclear localization signal. CTD52 induces a selective reorganization of splicing factors from discrete nuclear domains to the diffuse nucleoplasm, and significantly, it blocks the accumulation of spliced, but not unspliced, human β-globin transcripts. The extent of splicing factor disruption, and the degree of inhibition of splicing, are proportional to the number of heptapeptides added to the protein.

These results indicate a functional interaction between Pol II's CTD and pre-mRNA

splicing. Proteins containing the full length CTD disrupt splicing factor domains, and inhibit splicing *in vivo*. Thus, intact CTD proteins are probably too disruptive for targeting applications. Proteins containing different numbers of CTD heptapeptide repeats have been tested for their ability to target splicing factor domains and to inhibit splicing *in vivo*. The stepwise removal of heptapeptide repeats abolishes the inhibitory effect on splicing, but the short CTD peptides still appear to accumulate in the sites where RNA polymerase II, splicing factors and poly A+ RNAs are most concentrated. Therefore, short CTD-derived proteins should target bioactive molecules such as ribozymes to the sites where Pol II transcripts are synthesized and spliced.

Previous Determinations regarding CTD Peptides

As described in PCT/US95/15683, truncated CTD peptides retain a targeting activity that is distinct from the full length CTDs. The truncated CTD fusion proteins were examined for their ability to enter the nucleus and to localize to discrete domains. All CTDs tested could enter the nucleus without a NLS. Many of the CTD peptides are targeted to nuclear domains abutting or overlapping regions enriched with splicing proteins (i.e. ICGCs), but they also appear in the nucleoplasm surrounding the ICGCs.

For each construct, the B1C8 pattern was assessed in approximately 150 transfected nuclei. Transfected nuclei were scored as "unchanged ICGCs" or "reorganized ICGCs." Reorganized ICGCs fall into two categories: (i) total dispersal of B1C8; and (ii) shrinkage of ICGCs plus dispersal of B1C8 staining.

All CTD peptides (including CTD_1) bestow speckle localizing activity on indicator peptides.

A heptapeptide-like sequence located immediately upstream of the 'true' heptapeptide #1 must be removed to eliminate all speckle localizing activity from control peptide A. The targeting properties of serially truncated CTD peptides were different. Targeting activity increased as consensus heptapeptides were added to Flag-CTD_1, and reached a maximum at Flag-CTD_{26}, which is comprised of the upstream one-half of the CTD (that is, nearly all consensus repeats). The speckle-localizing activity declined as natural downstream variant repeats were added to the COOH terminus of Flag-CTD_{26}. The decline seems to be proportional to the number of variant heptads added (a "dose-dependent" phenomenon). As variant repeats were added the Flag-CTDs may become more competent to redistribute from the storage domains to the diffuse nucleoplasm where the genes are transcribed and spliced. Consistent with this idea, F-CTD_{52} has lower speckle-localizing ability than shorter CTD peptide, but it is fully competent to redistribute to enlarged speckle domains when the cells are treated with transcriptional inhibitors, similar to endogenous Pol II LS. A substantial fraction of endogenous Pol II LS resides in the diffuse nucleoplasm, and redistributes to the speckles in all transcriptionally inhibited cells. CTD_{26}, CTD_{13}, and CTD_6 target the Flag peptide to discrete domains better than longer CTDs.

Short CTDs, derived from the consensus-rich (upstream) half of the CTD, preferentially target indicator peptides to 'storage depots,' where they are essentially stranded, whereas the full length CTDs retain the ability to redistribute between the 'storage' sites and the sites where Pol II transcribes genes ("round trip ticket").

CTD peptides induce a dramatic reorganization of ICGs *in vivo*. Even the first three heptapeptide repeats (21 amino acids) can induce partial reorganization of the ICGCs *in vivo*.

Effect of Phosphorylation of CTDs

Pol II LS's CTD is either hyperphosphorylated or hypophosphorylated in vivo

All known phosphorylation sites are in Pol II LS's CTD. In mammalian cells, there are two major forms of Pol II LS (Fig 1B): "Pol IIo" is hyperphosphorylated predominantly on Ser/Thr residues at positions 2, 4, 5 and 7 of the CTD heptapeptide repeats, and migrates at approximately 240 kDa. "Pol IIa" is relatively hypophosphorylated and migrates at approximately 220 kDa. Because 241 of the 365 amino acid residues in the CTD are phosphorylatable (128 serines, 61 threonines and 52 tyrosines), there is potential for a vast array of differentially phosphorylated species of Pol II LS. Nevertheless, very few Pol II LS molecules migrate between 220 and 240 kDa *in vivo*; the majority of Pol II LS is either Pol IIo or Pol IIa. The implication is that intermediately phosphorylated Pol II LS species are rapidly converted to Pol IIo or Pol IIa.

Pol II LS's CTD is hyperphosphorylated at the onset of transcriptional elongation

CTD phosphorylation has been studied almost exclusively in *in vitro* transcription assays. Pol IIa is efficiently recruited to transcription initiation complex *in vitro*, and the elongation phase is heralded by phosphorylation of the CTD to yield Pol IIo. *In vivo*, paused polymerases are primarily Pol IIa, but they are converted to Pol IIo as they enter the elongation phase. The fate of Pol IIo molecules after the elongation phase is unknown, although studies demonstrate that Pol IIo can be stored. Therefore, hyperphosphorylation of

the CTD does not necessarily indicate that Pol II LS is engaged in transcriptional elongation. The level of Pol IIo remains unchanged throughout the cell cycle, including mitosis, when transcription is shut off. A subset of Pol IIo molecules is tightly sequestered in non-chromosomal locations inside and outside of the nucleus. Studies indicate that the CTD plays a pivotal role in Pol II LS's dynamic, cell cycle regulated redistribution between storage sites and the sites of pre-mRNA synthesis.

When expressed *in vivo*, isolated CTD proteins are phosphorylated, yielding two major forms analogous to Pol IIo and Pol IIa. Isolated CTD peptides are also hyperphosphorylated *in vivo*.

The data indicates that the CTD is a critical link between Pol II transcription complexes and the splicing machinery. RNA polymerase II and the splicing proteins may be stored in common or overlapping compartments, and may be coordinately recruited from these domains to the sites of pre-mRNA synthesis.

Flag-Tagged CTD_{52} is targeted to domains inhabited by endogenous Pol II LS. The CTD_{52} peptide mediates transcription-dependent and redistribution that coincides spatially and temporally with endogenous Pol II LS. As shown by immunofluorescence studies, F-CTd_{52} redistributes from a finely stippled distribution to a smaller number of enlarged speckle domains, which coincide with Pol IIo-enriched domains. Similar results were obtained with α-amanitin, which neither binds to nor induces dephosphorylation of the CTD. Similar results have also been obtained with β-Gal-CTD_{52}.

Rational design of CTD Conjugates

The properties of a repetitive polymer formed of multiple tandemly arranged heptapeptides will be determined by the number of heptapeptides; the ratio of consensus to variant heptapeptides; their relative arrangement in the polymer; and the phosphorylation state of each heptapeptide. Using this information, one can control the bioavailability of the pharmacological or bioactive agent being targeted to Pol II genes or to the associated splicing apparatus and/or the nucleus. For example, some synthetic "mini-CTDs" may be used to target the therapeutic agent preferentially to a 'storage' domains, with a limited ability to redistribute to the Pol II genes and active splicing machinery. Other CTD targeting modules may be constructed with features that allow them to more readily exit the 'storage' compartments (PRFs). The latter CTD modules would be more bioavailable to the Pol II transcription and splicing machinery.

The targeting modules may be constructed from two or more heptapeptide units with the structure: $YX_1PX_2X_3PX_4$, where Y is tyrosine, P is proline, X can be any amino acid, and X_1, X_2 and X_3 are most preferably serine or threonine, covalently conjugated to a bioactive molecule, most preferably an oligonucleotide, preferably via a linker consisting of one to two amino acids or a carbon chain of equivalent length attached using amide or disulfide chemistry to the tyrosine at the N-terminus of the peptide, leaving a free carboxyl.

In the naturally occurring mammalian Pol II LS CTD, the following heptapeptides can be used as building blocks for the CTD peptides:

YSPTSPS	YSPTSPN	YTPTSPN	YSPTSPA	YTPQSPS
YEPRSPGG	YSPTSPT	YSPTSPK	YTPTSPK	YSPTTPK
YSPTSPV	YSPTSPG	YSLTSPA	YTPSSPS	YSPSSPS
YTPTSPS	YSPSSPE	YTPQSPT	YSPSSPR	

Note: underlined residues vary from the consensus YSPTSPS.

**Key for single letter amino acid designations.

A=alanine	D=aspartic acid	Q=glutamine
I=isoleucine	R=arginine	C=cysteine
G=glycine	L=leucine	
N=asparagine	E=glutamic acid	H=histidine
K=lysine		
F=phenylalanine	F=phenylalanine	T-threonine
Y-tyrosine	F=phenylalanine	S=serine
W=tryptophan	V=valine	

The peptide can be phosphorylated to alter the association with certain molecules in the nucleus, such as proteins having a serine/arginine motif and Sm snRNPs. For example, it is demonstrated that phosphorylated CTD-derived peptides bind to these nuclear proteins associated with transcription and splicing. Accordingly, for delivery of molecules which are desired to be in close association with RNA, it may be desirable to phosphorylate the peptide, and for delivery of molecules which are not desired to be in close association with RNA, it may be desirable to leave the peptide unphosphorylated.

Labelling of Peptides

Although referred to herein as conjugates of bioactive molecules, unless otherwise stated, the term "bioactive" includes labels. The CTD-derived peptides can be directly or indirectly labelled with a detectable label to facilitate detection of the presence of the peptides by detection of the label. Various types of labels and methods of

labelling antibodies are well known to those skilled in the art. Several specific labels are set forth below.

For example, the peptide can be labelled directly or indirectly with a radiolabel such as, but not restricted to, ^{32}P, ^{3}H, ^{14}C, ^{35}S, ^{125}I, or ^{131}I. The radiolabel is generally attached by chemical modification. Detection of a label can be by methods such as scintillation counting, gamma ray spectrometry or autoradiography.

Fluorogens can also be used directly or indirectly to label the CTD peptides. Examples of fluorogens include fluorescein and derivatives, phycoerythrin, allo-phycocyanin, phycocyanin, rhodamine, Texas Red or other proprietary fluorogens. The fluorogens are generally attached by chemical modification and can be detected by a fluorescence detector.

The CTD peptide can alternatively be labelled directly or indirectly with a chromogen to provide an enzyme or antibody label. For example, the peptide can be biotinylated so that it can be utilized in a biotin-avidin reaction which may also be coupled to a label such as an enzyme or fluorogen. The peptide can alternatively be labelled with peroxidase, alkaline phosphatase or other enzymes giving a chromogenic or fluorogenic reaction upon addition of substrate. Additives such as 5-amino-2,3-dihydro-1,4-phthalazinedione (also known as Luminol™) (Sigma Chemical Company, St. Louis, MO) and rate enhancers such as p-hydroxybiphenyl (also known as p-phenylphenol) (Sigma Chemical Company, St. Louis, MO) can be used to amplify enzymes such as horseradish peroxidase through a luminescent reaction; and luminogeneic or fluorogenic dioxetane derivatives of enzyme substrates can also be used. Such labels can be

detected using enzyme-linked immunoassays (ELISA) or by detecting a color change with the aid of a spectrophotometer.

Bioactive Molecules

Bioactive molecules include proteins or peptides, sugars, and nucleic acid sequences, which can be ribozymes, external guide sequences for RNAase P, antisense, aptamers, triplex forming oligonucleotides, nucleosides, nucleotides, genes, cDNA, mRNA, or RNA.

As demonstrated by the data described herein, the RNA pol II and CTD peptides have been demonstrated to deliver molecules as large as beta-galactosidase, as well as the Flag-epitope, to the nucleus. Other examples of useful bioactive agents are polysaccharides, minerals, inorganic compounds and organic compounds.

Genes for use in human gene transfer are reviewed, for example, by Anderson, Science 260, 808 (1992), Miller, Nature 357, 455 (1992), Mulligan, Science 260, 926 (1993), and Crystal, Science 270, 404 (1995). Ribozymes are reviewed, for example, by Sullenger and Cech, Science 262, 1566-1569 (1993), Barinaga, Science 262, 1512-1514 (1993), Yu, et al., Proc. Natl. Acad. Sci. USA 90, 6340-6344 (1993), and external guide sequences for targeted cleavage by ribonuclease P is described by Altman, Proc. Natl. Acad. Sci. USA 90(23), 10898-10900 (1993), Liu and Altman, Genes & Devel. 9(4), 471-480 (1995), Yuan, et al., Proc. Natl. Acad. Sci. USA 89(17), 8006-8010 (1992), and Altman, et al., FASEB J. 7(1), 7-14 (1993). Antisense oligonucleotides are described by Herrmann, J. Molec. Med. 73, 157-163 (1995), De Clercq, Clin. Microbiol. Rev. 8, 200-239 (1995), Wagner, Nature 372, 333-335 (1994), Crooke Antisense Res. Develop. 3, 1-2 (1993), Agrawal and Akhtar, Trends in

Biotechnol. 13, 197-199 (1995), Agrawal, et al., Clin. Pharmacokinetics 28, 7-16 (1995), Agrawal, et al., Current Opinion in Biotechnol. 6, 9-12 (1995), Temsamani, et al., Antisense Res. & Devel. 4, 35-42 (1994), and Zamecnik, et al., Nucleic Acids Symp. Series 24, 127-131 (1991). The Peptides also can be used to deliver procaryotic and eucaryotic cells, e.g., bacteria, yeast, and mammalian cells, including human cells, and components thereof, such as cell walls, and conjugates of cellular components.

Peptide can also be used to deliver water soluble or water insoluble drugs such as anesthetics, chemotherapeutic agents, immunosuppressive agents, steroids, antibiotics, antivirals, antifungals, antiinflammatories, and anti-parasitic drugs.

Imaging agents also may be attached to peptide, including metals, radioactive isotopes, radioopaque agents, fluorescent dyes, and radiolucent agents. Radioisotopes and radioopaque agents include gallium, technetium, indium, strontium, iodine, barium, and phosphorus.

Coupling of Bioactive Agents

The bioactive molecules are preferably covalently coupled to the CTD-derived peptides using standard chemistry, such as amide or sulfide coupling chemistries. These methods are known to those skilled in the art for coupling of proteins, nucleic acids, polysaccharides, and combinations thereof.

One useful protocol involves the "activation" of hydroxyl groups on the Peptide carbonyldiimidazole (CDI) in aprotic solvents such as DMSO, acetone, or THF. CDI forms an imidazolyl carbamate complex with the hydroxyl group which may be displaced by binding the free amino group of a

bioactive ligand such as a protein. The reaction is an N-nucleophilic substitution and results in a stable N-alkylcarbamate linkage of the ligand to the Peptide. The resulting ligand-Peptide complex is stable and resists hydrolysis for extended periods of time.

Another coupling method involves the use of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDAC) or "water-soluble CDI" in conjunction with N-hydroxylsulfosuccinimide (sulfo NHS) to couple the exposed carboxylic groups of the Peptide to the free amino groups of bioactive ligands. EDAC and sulfo-NHS form an activated ester with the carboxylic acid groups of the Peptide which react with the amine end of a ligand to form a Peptide bond. The resulting peptide bond is resistant to hydrolysis. The use of sulfo-NHS in the reaction increases the efficiency of the EDAC coupling by a factor of ten-fold and provides for exceptionally gentle conditions that ensure the viability of the ligand-peptide complex. These protocols permit the activation of either hydroxyl or carboxyl groups on the peptide, and attachment of the desired bioactive ligand.

A useful coupling procedure for attaching ligands with free hydroxyl and carboxyl groups to the Peptide involves the use of the cross-linking agent, divinylsulfone. This method is useful for attaching sugars or other hydroxylic compounds to hydroxyl groups on the peptide. The activation involves the reaction of divinylsulfone with the hydroxyl groups of the peptide to a vinylsulfonyl ethyl ether. The vinyl groups will couple to alcohols, phenols and amines. Activation and coupling take place at pH 11. The linkage is stable in the pH range from 1-8 and is suitable for transit through the intestine. Any suitable

coupling method known to those skilled in the art may be used to couple bioactive ligands to the Peptide.

The bioactive agent can be covalently coupled to peptide either directly or indirectly using a linker molecule. Linker molecules will typically be used when additional flexibility or space is needed between the peptide and the therapeutic compound. Any suitable molecule that can be coupled to both protein and a bioactive agent can be used as a linked.. Exemplary linkers are peptides or molecules with straight carbon chains.

Peptides modified to increase *in vivo* half-lives

The peptides can be prepared by recombinant techniques and expression in an appropriate host systems, isolated from natural sources as described above, or prepared by synthetic means. These methods are known to those skilled in the art. An example is the solid phase synthesis described by J. Merrifield, 1964 *J. Am. Chem. Soc.* 85, 2149, used in U.S. Patent No. 4,792,525, and described in U.S. Patent No. 4,244,946, wherein a protected alpha-amino acid is coupled to a suitable resin, to initiate synthesis of a peptide starting from the C-terminus of the peptide. Other methods of synthesis are described in U.S. Patent No. 4,305,872 and 4,316,891.

Peptides containing cyclopropyl amino acids, or amino acids derivatized in a similar fashion, can also be used. These peptides retain their original activity but have increased half-lives *in vivo*. Methods known for modifying amino acids, and their use, are known to those skilled in the art, for example, as described in U.S. Patent No. 4,629,784 to Stammer.

The peptide can also be administered as a pharmaceutically acceptable acid- or base- addition salt, formed by reaction with inorganic acids such as hydrochloric acid, hydrobromic acid, perchloric acid, nitric acid, thiocyanic acid, sulfuric acid, and phosphoric acid, and organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, pyruvic acid, oxalic acid, malonic acid, succinic acid, maleic acid, and fumaric acid, or by reaction with an inorganic base such as sodium hydroxide, ammonium hydroxide, potassium hydroxide, and organic bases such as mono-, di-, trialkyl and aryl amines and substituted ethanolamines.

Therapeutic Applications

Since the CTD peptides accumulate precisely in discrete sites inhabited by RNA polymerase II and the spliceosomes, they should have significant applications in genetic therapy technologies and in targeting of drugs that act in the nucleus. CTD peptides may be attached to antisense oligonucleotides, catalytic RNAs and transgenes, as described above, to deliver and concentrate the nucleotide sequences in the nuclear compartment where the pre-mRNAs are synthesized and processed. The CTD peptides may minimize intranuclear sequestration of therapeutic polynucleotides.

In the preferred embodiment, the peptides are coupled to bioactive agent, then administered to a patient in need thereof. Those skilled in the art will know how much and in what formulation the peptide-drug conjugate is administered. Formulations for intravenous administration include saline and phosphate buffered solution as well as liposomes and other microparticulates which increase bioavailability and can be used for targeting to specific tissues; formulations for

topical, transmucosal, and aerosol formulations are similarly available, for example, as described in Goodman and Gilmans. In some cases, the peptide-drug conjugates may be administerable orally, preferably in an enteric coating or microencapsulated to enhance uptake and bioavailability.

The present invention will be further understood by reference to the following non-limiting examples demonstrating that the highly conserved and repetitive CTD links splicing components to a key subunit of RNA polymerase II, thereby helping to coordinate the processes of transcription and splicing.

Recently, a fraction of Pol IIo was immunolocalized in 20-50 discrete nuclear domains ("speckles"), which are enriched with SR splicing proteins and Sm snRNPs (Bregman et al., 1995; Blencowe et al., 1996). In addition, Pol IIo, SR proteins and Sm snRNPs become sequestered in dot-like non-chromosomal domains during mitosis, when transcription is inactive (Warren et al., J. Cell Sci. 103:381-388 (1994); Bregman et al., J. Cell Sci. 10: 387-396 (1994)). These immunolocalization experiments revealed Pol IIo molecules in the same non-chromosomal location as certain splicing factors, but it was the preceding study which showed for the first time that splicing factors are associated with Pol IIo in the absence of pre-mRNA, and at times when the polymerase is not engaged in transcription. The latter findings, together with the observation that anti-CTD phosphoepitope-specific mAbs H5 and H14 can release Pol IIo from the splicing factors *in vitro*, strongly imply that Pol IIo's association with the splicing factors is mediated by the hyperphosphorylated CTD. Indeed, the results of

the latter study prompted a study to determine whether the CTD interacts with the pre-mRNA splicing process *in vivo*.

Overexpression of CTD-derived proteins results in the dispersal of Sm snRNPs and SR splicing factors from a speckled pattern to a diffuse nucleoplasmic distribution. This property is selective, since other types of nuclear domains remain intact. CTD-derived proteins block the accumulation of spliced, but not unspliced, human β-globin transcripts *in vivo*. The stepwise addition of heptapeptide repeats to a fusion protein potentiates its ability to disrupt the splicing factor domains, and to inhibit splicing *in vivo*. These results, in conjunction with the preceding study, strongly suggest that the highly conserved and repetitive CTD links splicing components to a key subunit of RNA polymerase II, thereby helping to coordinate the processes of transcription and splicing.

The following materials and methods were used in these studies.

Plasmids expressing Flag-tagged CTD-derived proteins. Epitope-tagged CTD expression plasmids were created using standard techniques (Sambrook et al., Molecular Cloning-A Laboratory Manual. 2nd Ed. (Cold Spring Harbor Press, Cold Spring Harbor, NY. 1989). Full length CTD coding sequences were obtained from a human Pol II LS cDNA isolated and sequenced by L. Du. The Pol II LS cDNA was authenticated by comparison to EMBL sequence X63564 (Wintzerith et al., Nuc. Acids. Res. 20:910 1992). A 2.1 kb BamHI fragment containing the C-terminal domain plus 146 bp of 3'-untranslated mRNA was subcloned into the BamHI site of pcDNA3AB, an expression vector derived from pcDNA3 (Invitrogen). pcDNA3AB has a Flag^R epitope

(AspTyrLysAspAspAspAspLys; Kodak) immediately upstream of the multiple cloning site. The full length Flag-tagged CTD expression plasmid is termed "pF-CTD52" to indicate the presence of 52 heptapeptide repeats. pF-CTD52 is predicted to express a fusion protein comprised of an N-terminal Flag peptide^{R} attached to 636 amino acids derived from the C-terminus of human Pol II LS. The latter segment includes residues 1335-1588 (immediately upstream of the CTD) and residues 1589-1970, which contain 52 tandemly repeated heptapeptides. pF-CTD32 (analogous to pF-CTD52, but lacking heptapeptides 33-52) was derived from a BamHI/EcoRI cDNA clone isolated from a human fetal liver library (Stratagene); sequence analysis revealed this fragment to be truncated within the 32nd repeat of the CTD coding sequence. pF-CTD26 (identical to pF-CTD52, but lacking heptapeptides 27-52) was made as follows: the 2.1 kb BamHI fragment of human Pol II LS cDNA (from nt 4001 to nt 6059 of coding region) was digested with SpeI. The resulting 1.3 kb BamHI-SpeI fragment was subcloned into the BamHI-XbaI sites of pcDNA3AB. pF-CTD13, pF-CTD6, pF-CTD3 and pF-CTD1 were generated by PCR mutagenesis. The forward primer for each of these reactions was p4204U (5'AAGAGGTGGTGGACAAGATGGATG-3'), an oligonucleotide that hybridizes to a 24 nucleotide sequence 183-160 base pairs upstream of the BamHI site located at nucleotides 4001-4006 within coding portion of the human Pol II LS cDNA. Reverse primers included: p5394 (5'-GCGAATTCGCTGGGAGAGGTGGGCGAATAGCT-3') for pF-CTD13; p5264 (5'-GCGAATTCGGACTGGTTGGAGAATAGGATGGA-3') for pF-CTD6; p5205 (5'-CGAATTCAAGAGGGACTCTGGGGTGTGTAGCC-3') for pF-CTD3, and p1CTD (5'-GCGAATTCAGCTTGGACTAGTGGGTGAGTAGCTGGGA

GACATGGCGCCACCTGGTGA-3') for pF-CTD1. The PCR products were digested with EcoRI (encoded in downstream primers) and BamHI (present in Pol II cDNA 183-160 nucleotides downstream of the upstream primer). The PCR products were subcloned into the BamHI/EcoRI sites of pcDNA3AB.

One control plasmid, pF-CTDless.3, expresses the Flag-tagged N-terminal 282 amino acids of Pol II LS. This segment was generated by PCR amplification, using human Pol II LS cDNA as the template. The oligos for this reaction were: p337U (5'-GCGAATTCGGCTTTTTGTAGTGAGGTTTG-3') and p1209L (5'GCGAATTCGTCAGCCAGTTTGTGAGTCAGGTC-3'). The amplified segment of DNA was digested with EcoRI and subcloned into the EcoRI site of pcDNA2AB. Another control plasmid, pF-CTDless.1, expresses a Flag-tagged approximately 25 kDa segment of Pol II immediately upstream of the CTD. This control sequence corresponds to a 714 bp BamHI-SmaI fragment derived from the Pol II LS cDNA, which was subcloned into the BamHI-EcoRV sites of pcDNA3AB. A third control, pF-CTDless.2, expresses a Flag-tagged approximately 22 kDa segment of Pol II LS, which contains a 500 bp fragment immediately downstream from the BamHI site. This fragment was generated by PCR amplification using oligos p4204U and p4869L (5'-CGAATTCAGCCGGTGGGTCCAGCAGC-3'). The PCR product was digested with BamHI and EcoRI and subcloned into pcDNA3AB. This F-CTDless.2 protein is similar to F-CTDless.1, but lacks a heptapeptide-like sequence (MFFGSAPSPMGGISPAMTPWNQGATPAYGAWSPSVGSGMTPGAAGFSPSA ASDASGFSPGYSPAWSPTPGSPGSPGPSSPYIPSPGGA), which precedes the CTD.

Plasmids expressing βGalactosidase-linked CTD proteins. The βGal-CTD fusion constructs were made

as follows: First, the stop codon at the end of the βGalactosidase gene was replaced with restriction sites EcoNI, BamHI and SalI, which were recombinantly added to the C-terminus. This PCR reaction utilized pSVβ (Promega) as a template, and two primers: a downstream adapter oligo which contains a SalI site (pMCS, 5'-GCGTCGACTCTAGAATTCGCGGATCCTCCTGAAGGTTTTTGACACCAGACC AACTGG-3') and an internal oligo p3047 (5'-GGATTGGTGGCGACGACTCCTGGA-3'). The 130 bp PCR product was digested with Esp I and Sal I and inserted back into pSVβ that had been cut with Esp I and Sal I to make pSVβMCS. Next, the βGal coding sequence was excised from pSVβMCS with SmaI and BamHI, and subcloned into pcDNA3 that has been cut with Hind III, filled in with Klenow and cut with BamHI. The resulting vector, pcDNAβGal, preserves all the cloning sites downstream of BamHI from pcDNA3. Finally, CTD-26, CTD-32 and CTD-52 fragments with BamHI-EcoRI ends were subcloned into the corresponding sites of pcDNAβGal to generate the βGal-CTD series (Figure 1 shows only pβGal-CTD52, pβGal-CTDless and βGal).

Plasmids expressing human β-globin genes and recombinant CTD-derived proteins. Plasmids that co-express Flag-tagged CTD-derived proteins and human β-globin genes are generically termed "pF-CTD_xβ-globin [+/-]," where "F" refers to the Flag peptide coding sequence, "CTD" refers to the sequence of the CTD-derived protein , "x" refers to the number of heptapeptide repeats, "β-globin" refers to the β-globin gene, and the "[+]" and "[-]" signs designate the relative orientation of the two transcription units.

The plasmids were constructed as follows: A 2.7 kb HindIII-FspI fragment containing the 2.3 kb human β-globin gene plus an SV40 enhancer element

was excised from pUCβ128SV (Caceres et al., Science 265:1706-1709 (1994)), filled in with Klenow fragment of DNA polymerase I, and subcloned into CTD expression plasmids pF-CTD1, pF-CTD6, pF-CTD13 and pF-CTD52, each of which had been digested with EcoRV. For controls, β-globin genes were subcloned into pF-CTDless.1, pF-CTDless.3 and pβgal.

Antibodies. Monoclonal antibodies (mAbs) H5 and H14 are described by Bregman, et al. (1995). mAb Y12 (Lerner, et al. Proc. Natl. Acad. Sci. USA. 78:2737-2741 (1981); Pinto, et al. Proc. Natl. Acad. Sci. USA. 86:8742-8746 (1989)) recognizes the Sm snRNP B/B' and D, as well as a 70 kD proteolytic fragment of intron binding protein (IBP). B1C8 is described by Kim et al., J. Cell Biol. 1997). MAb M2 (Kodak) is an IgG that binds to the Flag^R peptide, AspTyrLysAspAspAspAspLys. MAb anti-βGal is an IgG that binds to βGalactosidase (Promega). PAb anti-βGal is a polyclonal antibody that binds to βGalactosidase (Cappel). MAb 138 is an IgG directed against ND55, a protein in N10 domains (Ascoli and Maul, J. Cell Biol. 112:785-95 1991). Anti-coilin is a rabbit antiserum directed against p80 coilin (Andrade et al., Proc. Natl. Acad. Sci. 90:1947-1951 1993).

Cell Culture and transient plasmid transfections. Cells were maintained as described by Bregman, et al., (1994). For *in vivo* splicing assays, 10^6 HeLa cells were seeded in a 60 mm petri dish and transfected with each of the plasmids (5 μg) using 45 μl of Lipofectamine^R.

SDS-PAGE and immunoblotting. HeLa cell nuclei were extracted into ice cold 50 mM Trix-HCl, 250 mM NaCl, 1% NaCl, 1% Triton X-100, 1 mM PMSF, 0.2 mM $NaVO_4$, 5 mM β-glycerophosphate, pH 7.4 (T-buffer), insoluble material removed by centrifugation, and the supernatant immunoprecipitated. Nuclear

extracts were incubated with 50 microliters of staphylococcal protein G coupled-agarose beads, pre-charged with an excess of antibody. AFter incubation for 4 hours at 4°C, the beads were washed three times with ice cold T-buffer, boiled in SDS sample buffer and subjected to SDS-PAGE and immunoblotting. Immunoblotting was performed as described by Bregman, et al. (1995).

Immunofluorescence microscopy and image analysis. Immunofluorescence microscopy is performed as described by Bregman, et al. (1994). Briefly, cells grown on coverslips were fixed with 1.75% paraformaldehyde and permeabilized with 0.5% Triton X-100. Nonspecific binding sites were blocked by incubating coverslips wtih 4% bovine serum albumin in Dulbecco's phosphate buffered saline (DPBS) followed by incubation with specific antibodies diluted in DPBS containing 0.5% BSA. Specific antibody binding was visualized by incubating washed coverslips with fluorescein conjugated anti-IgM, or rhodamine conjugated anti-IgG diluted in PBS with 0.5% BSA (Vector Laboratories). To visualize chromosomes, cells were stained with 4',6-diamidino-2-phenylindole (DAPI) at 5 μg/ml as described by Warren, et al. (1992). Images were captured using a Photometrics CH250 CCD camera (CE200A/LC200 liquid cooling), mounted on an Axioskop (Zeiss). Image analysis software included NIH-Image, Registration v.1.1d2[R] and Adobe Photoshop 2.0[R].

Quantitative RT-PCR and RNase protection assays. Quantitative RT-PCR was carried out as follows: Total RNA was prepared from HeLa cells 1 to 2 days after transfection using UltraSpec RNA[R] (Biotecx), and digested with RNase-free RQ1 DNase (Promega) to remove contaminating DNA. The RNA was phenol extracted, ethanol precipitated and dissolved in

water. A reverse primer (449 nucleotides downstream from the HindIII site) that hybridizes to the second exon of the β-globin gene, 5'-CAGGAGTGGACAGATCCC-3', was used for reverse transcription, followed by 10, 12, and 14 cycles of PCR amplification using a forward oligo 5'-TCAAACAGACACCATGGTGCACCTGACT-3' which hybridizes to exon 1 of β-globin (167 nucleotides downstream from the HindIII site).

The RNase protection assay was carried out using the RPA II Ribonuclease Protection Kit (Ambion) according to the manufacture's procedures. Briefly, a HindIII-BamHI fragment containing exon 1, intron 1 and most of exon 2 of human β-globin was subcloned into the corresponding sites of BlueScript-SKII (Strategene) and digested with BbvII located within intron 1 to yield a linearized template. A complementary RNA probe was synthesized *in vitro* with T3 RNA polymerase (New England Biolabs) in the presence of 10 units of RNasin (Promega), 0.5 mM of ATP, GTP, UTP, 3 μM of CTP and 100 μCi [α-^{32}P]CTP, which yielded an internally radiolabeled 343 nt fragment covering exon 2 and the 3' half of intron 1. The probe was purified on a 4.5% denaturing polyacrylamide gel and hybridized to total RNA prepared from transfected HeLa cells at 45 °C overnight. Hybridization mixtures were digested with RNase A/T1, precipitated, solubilized with gel loading buffer and separated on a 4% denaturing polyacrylamide gel. The dried gel was exposed to hyperfilm (Amersham) overnight and scanned into a digital image using ScanJet (Hewlett Packard) and analyzed using NIH Image^{R} software.

Example 1: CTD-derived Peptides localize in the nucleus.

Pol IIo's association with pre-mRNA splicing factors was hypothesized to be mediated by the CTD. To test this hypothesis, CTD-derived sequences which lack the catalytic and DNA binding regions of the Pol II LS were tested to see if they could target indicator proteins to speckle domains. For this purpose, the Flag peptide (Flag) or βGalactosidase (βGal) sequences were recombinantly added to the N-terminus of the CTD-containing proteins. The resulting fusion proteins were transiently expressed and immunolocalized in CV1 cells. Next, whether the CTD-derived fusion proteins interfere with splicing *in vivo* was determined. For this purpose, human β-globin pre-mRNAs and CTD-derived proteins were co-expressed in HeLa cells, and the efficiency of β-globin splicing *in vivo* quantitated.

Plasmid vectors that express a variety of CTD-containing fusion proteins were constructed as shown in Figure 1. The expression and intracellular distribution of each fusion protein has been documented by immunoblotting, immunoprecipitation and immunostaining with antibodies directed at the Flag epitope or βGal. The minimum number of heptapeptide repeats required to achieve certain biological effects were determined. The CTD-containing fusion proteins were unidirectionally truncated from the C-terminus, giving rise to a nested set of proteins containing heptapeptides 1-52 (F-CTD52 and βGal-CTD52); 1-32 (F-CTD32 and βGal-CTD32); 1-26 (F-CTD26 and βGal-CTD26); 1-13 (F-CTD13); 1-6 (F-CTD6); 1-3 (F-CTD3), or only the first heptapeptide (F-CTD1) (Figure 1). Several control proteins were

used: (i) F-CTDless.1; (ii) F-CTDless.2; (iii) F-CTDless.3; (iv) βGal-CTDless and (v) βGal.

The CTD-derived fusion proteins must gain access to the nucleus to interact with splicing factors. At the beginning of the study, each fusion protein was immunolocalized in CV1 or HeLa cells to confirm that the experimental approach meets this requirement. Plasmids expressing each of the 13 proteins illustrated in Figure 1 were transfected into cells. Two days later, the cells were fixed and double immunostained with: (i) an antibody directed at the indicator portion of the fusion protein (anti-Flag or anti-βGal), and (ii) an antibody directed at the CTD portion of the fusion protein (mAb H5 or mAb H14).

In a representative experiment, CV1 cells were transfected with pF-CTD52. Two days later the cells were fixed and double immunostained with anti-Flag^R mAb M2 and anti-CTD mAb H14. Anti-Flag staining was visualized by rhodamine and mAb H14 staining was visualized by FITC. F-CTD52 is distributed almost exclusively in the nucleus, even though it lacks a conventional nuclear localization signal. In addition, the untransfected cell nucleus is weakly immunostained by mAb H14, whereas the transfected cell nucleus is intensely labeled. The F-CTD52 protein is present in the diffuse nucleoplasm, but it is most concentrated in approximately 50 discrete, non-nucleolar sites. In addition, the transfected cell nucleus is much more intensely stained by mAb H14 than the untransfected cell nucleus The nuclear "dots" are also intensely stained by mAb H14 and mAb H5 antibodies, both of which recognize CTD phosphoepitopes. These results indicate that F-CTD52 accumulates in the nucleus, and suggest that CTD heptapeptides on the F-CTD52 protein are phosphorylated *in vivo*. All of the

CTD-derived and control proteins illustrated in Figure 1 are expressed and enter the nucleus.

Example 2: The CTD-derived proteins are phosphorylated *in vivo*.

All observations indicating an association between Pol II LS and splicing factors suggest a mechanism involving a hyperphosphorylated CTD (Bregman et al., 1995; Kim et al., 1997; Blencowe et al., 1996). Therefore, if the CTD-derived proteins are expected to interact with splicing factors in the nucleus, they probably need to be phosphorylated *in vivo*. The immunolocalization studies described above suggest strongly that the CTD-derived fusion proteins are phosphorylated *in vivo*. To confirm this impression, and to establish the electrophoretic mobility of each CTD-derived protein, whole cell extracts were prepared from cells transfected with each plasmid in Figure 1. The samples were subjected to 5-15% gradient SDS-PAGE and immunoblotted with: (i) mAbs directed against CTD-specific phosphoepitopes (H5 or H14); or (ii) mAbs directed at the indicator part of the protein (Flag or βGal).

In vivo phosphorylation and nuclear localization of CTD-derived fusion proteins was demonstrated as follows. CV1 cells were transfected with plasmids. Two days later, the cells were lysed in SDS sample buffer, subjected to 5-15% gradient SDS-PAGE and immunoblotted with antibodies: MAb M2, directed against the Flag® epitope, anti-βGal directed against β-Galactosidase and mAbs H5 and H14, directed against CTD phosphoepitopes. A 160 kD SR-related family splicing factor was immunolocalized with mAb B1C8.

An analysis of the Flag-tagged proteins showed that mAbs H14 and H5 blot an approximately 240 kDa

protein corresponding to endogenous Pol IIo in all of the extracts. In cells transfected with the pFCTD series of plasmids, mAbs H5 and H14 14 blot a nested set of fusion proteins. In this experiment, mAb H5 immunoblots F-CTD26, F-CTD32 and F-CTD52, and mAb H14 immunoblots pF-CTD6, pF-CTD13, pF-CTD26, pF-CTD32 and pF-CTD52. As expected, the stepwise removal of heptapeptide repeats incrementally increases the electrophoretic mobility of the proteins. However, the apparent MW of each fusion protein significantly exceeds its predicted size. For example, F-CTD52 migrates as a 120/130 kDa doublet, even though it has a predicted MW of approximately 66 kDa. Repeated immunoblotting experiments reveal that many of the CTD-derived proteins migrate as closely spaced doublets.

The anomalous SDS-PAGE mobilities of the CTD-derived proteins, and the observation that alkaline phosphatase treatment of the filters abolishes mAb H14 and H5 immunoreactivity, indicate that the CTD-derived proteins are phosphorylated. Together with the previous studies showing that mAbs H5 and H14 recognize distinct phosphoepitopes on the CTD of native Pol II, these data indicate that the phosphorylation sites are within the CTD portion of the fusion proteins.

Some of the Flag-tagged CTD proteins are immunoblotted weakly, or not at all, by anti-Flag mAb M2. However, all of the Flag tagged CTD-derived proteins are expressed in HeLa or CV1 cells, since anti-Flag mAb M2 stains the nucleus in cells transfected by pF-CTD1, pF-CTD3, pF-CTD6, pF-CTD13, pF-CTD26, pF-CTD32, and pF-CTD52.

Some short CTD-derived proteins are not immunoblotted by mAbs H5 and H14. Nevertheless, it is believed that all of the FCTD proteins are

phosphorylated in the cell, as indicated by enhanced mAb H14 immunostaining of transfected cell nuclei. The inability of mAb H5 to immunoblot F-CTD1, F-CTD3, F-CTD6 and F-CTD13, and the inability of mAb H14 to immunoblot F-CTD1 and F-CTD3, may be explained by three factors: (i) Transfection efficiencies vary widely from experiment to experiment and from plasmid to plasmid. (ii) Fusion proteins with only a few heptapeptides have fewer potential phosphorylation sites, and hence fewer mAb H5 and H14 binding sites, than proteins with long CTD segments (e.g. F-CTD52 has approximately 50-fold more phosphorylation sites than F-CTD1). (iii) Finally, it is possible that downstream heptapeptides are better kinase substrates than upstream heptapeptides. In this regard, it is interesting to note that repeats 1-3 diverge from the YSPTSPS consensus sequence more than other repeats in the CTD.

An immunoblot of selected βGal linked CTD proteins shows that mAb H14 immunoblots an approximately 240 kDa protein corresponding to endogenous Pol IIo. MAb H14 also immunoblots βGalCTD fusion proteins in cells transfected with pβGal-CTD26, pβGal-CTD32 and pβGal-CTD52. Hyperphosphorylated βGal-CTD52 co-migrates with Pol IIo at approximately 240 kDa; however, more rapidly migrating species are observed in some experiments. Finally, immunoblotting with an antibody directed at βGal reveals the expected stepwise increase in the PAGE mobility of these proteins.

Example 3: Expression of F-CTD52 or βGal-CTD52 induces the SR-related splicing factor B1C8 to redistribute from discrete domains to a diffuse nucleoplasmic pattern.

The F-CTD52 protein is phosphorylated on CTD epitopes, and it enters the nucleus where it is frequently, but not always, observed in discrete nuclear "dots". One possible explanation for this distribution is that CTD52 targets the Flag peptide to splicing factor domains, perhaps reflecting its ability to associate with Sm snRNPs and SR family splicing proteins, which are most concentrated in the "speckles." To further explore this idea, it was examined whether the F-CTD52 containing "dots" overlap or co-localize with speckle domains. CV1 cells were transfected with pF-CTD52, and two days later the cells were double immunostained with anti-Flag mAb M2 (IgG) and mAb B1C8 (IgM), which recognizes an SR-related splicing protein in the speckle domains. CTD-derived fusion proteins and control proteins were immunolocalized with anti-Flag^R mAb M2 or anti-βGal. A 160 kD SR-related family splicing factor (Blencowe et al., 1995) was immunolocalized with mAb B1C8. ND55 was immunolocalized with mAb 138 (Ascoli and Maul, 1991).

This experiment yielded a striking and unexpected result: the B1C8 splicing factor is distributed in a speckled pattern in untransfected cell nuclei, but it is distributed in a nearly uniform, diffuse nuclear pattern in every cell expressing the F-CTD52 protein. Control proteins such as F-CTDless.1 accumulate in the nucleus, but they have little effect on the distribution of B1C8.

Next it was determined if CTD is responsible for the redistribution of B1C8. For this purpose, CV1 cells were transfected with pβGal-CTD52, and the cells were double immunostained with mAb B1C8 and anti-βGal. Again, B1C8 has a speckled distribution in control cells, but it has a diffuse nuclear distribution in cells expressing βGal-CTD52. B1C8 remains in a speckled distribution in nuclei expressing similar levels of a control protein, βGal-CTDless.

It was also determined whether F-CTD52 alters the distribution of proteins located in other types of nuclear domains. ND55 (55 kDa) is one of several proteins localized in approximately 10 highly circumscribed nuclear dots, referred to as "N10 domains" or "PML bodies" (Ascoli and Maul, 1991). N10 domains are dynamic structures. For example, the number of N10 domains increases after growth factor stimulation, and they disassemble following virus infection (Maul and Everett, J. Gen. Virol. 75:1223-33 (1994); Terris et al., Cancer Res. 55:1590-1597 (1995). Several proteins in N10 domains have been identified, but none appear to have a role in pre-mRNA splicing. Cells were transfected with pF-CTD52 and double immunostained with anti-ND55 mAb 138 (IgM) and anti-Flag mAb M2 (IgG). The results indicate that F-CTD52 does not alter the distribution of ND55, which remains exclusively in the N10 domains.

These results are consistent with the idea that the CTD interacts with splicing factors in the speckles.

Example 4: Addition of heptapeptide repeats to the fusion protein leads to an incremental disruption of B1C8 speckles.

The next goal was to determine how many heptapeptide repeats are required to induce the redistribution of B1C8. Therefore, a "heptapeptide titration" experiment was performed: CV1 cells were transfected with a nested set of Flag-tagged CTD-derived proteins: pF-CTD26, pF-CTD13, pF-CTD6, pF-CTD3 and pF-CTD1. Two days later, the cells were fixed and double immunostained with anti-Flag mAb M2 (IgG) and mAb B1C8 (IgM).

Immunostaining of F-CTD26 mAb with M2 reveals four transfected cell nuclei. Note that mAb M2 staining is almost exclusively intranuclear, and the level of FCTD26 expression varies widely among the four cells. Diffuse mAb M2-immunoreactivity is observed in all four nuclei, but two nuclei also contain discrete dots harboring the F-CTD26 protein. The nucleus expressing the highest level of F-CTD26 has a completely dispersed pattern of B1C8 staining. The nucleus expressing the second highest level of F-CTD26 has a nearly complete dispersal of B1C8 staining. The two nuclei expressing low levels of F-CTD6 have a partial dispersal of the B1C8 staining pattern as indicated by the multiple diminutive B1C8 speckles. Finally, the two untransfected nuclei each contain approximately 20 prominent B1C8-speckles. These results indicate that the upstream half of the CTD retains the ability to disrupt the distribution of B1C8, and the degree of B1C8 disruption is proportional to the level of CTD-derived protein in the nucleus. Similar results were obtained with F-CTD32.

Immunostaining of F-CTD13 with anti-Flag mAb M2 reveals a transfected cell nucleus and an untransfected cell nucleus. MAb M2 staining is almost exclusively intranuclear; the F-CTD13 protein is distributed in approximately 75 discrete dots, as well as the diffuse nucleoplasm. The nucleus expressing F-CTD13 has a dispersed pattern of B1C8 staining and the control nucleus has a typical speckled pattern. Thus, removal of 75% of the heptapeptides from the CTD does not abolish the B1C8-disrupting property of the fusion protein.

Three representative nuclei expressing low, medium and high levels of the F-CTD3 protein show mAb M2 staining which is almost exclusively intranuclear, and the distribution of F-CTD3 is diffuse with a few discrete dots. The nucleus expressing a low level of F-CTD3 retains a prominent speckled pattern of B1C8 staining. Nuclei expressing higher levels of F-CTD3 protein have a partial disruption of B1C8 staining, as indicated by diminutive speckles. Partial disruption of B1C8 stained speckles is observed in a nucleus expressing F-CTD6. The transfected nucleus has diminutive B1C8 speckles.

A representative cell nucleus transfected with F-CTD1 reveals mAb M2 staining in a diffuse and punctate distribution. Most nuclei expressing the F-CTD1 protein have prominent B1C8 containing speckles. When the anti-B1C8 and anti-Flag images are merged, one observes a close spatial relationship between the B1C8-speckles and F-CTD1 dots. Close examination of a nucleus expressing F-CTD6 reveals a similar phenomenon. Many of the overexpressed CTD proteins form discrete dots, and in nuclei containing intact B1C8 speckles the CTD-rich dots do not coincide with the speckles. Quantitative image analysis is needed to determine

whether the CTD-rich dots are organized randomly with respect to the B1C8 speckles, or whether they reproducibly form at the periphery of the speckles.

The effect of CTD length (i.e. number of heptapeptide repeats) on B1C8-speckles was quantitated as follows: CV1 cells were transfected with each of the Flag-tagged CTD-derived plasmids in Figure 1. The cells were fixed and double stained with anti-Flag mAb M2 and a mAb directed against B1C8. The pattern of B1C8 staining in each transfected cell nucleus was scored as "intact" (20-50 prominent speckles) or "disrupted" (diffuse pattern or diminutive speckles). Multiple sets of experiments were conducted, and 150-250 nuclei were scored for each plasmid.

The scoring results are presented in Figure 2. Intact B1C8 speckles were observed in greater than 90% of control (untransfected) nuclei (Figure 2, light gray bar). Intact speckles were observed in approximately 76% of nuclei expressing a control protein, F-CTDless.1 which contains a heptapeptide-like sequence on its C-terminus, which was derived from the region upstream of the CTD. These heptapeptide-like sequences are deleted in F-CTDless.2, and interestingly, intact speckles were observed in approximately 86% of nuclei expressing this control protein. Expression of F-CTDless.3, which has no heptapeptide-like sequences, does not reduce the frequency of intact B1C8 speckles. Intact B1C8 speckles were observed in approximately 70% of cell nuclei expressing F-CTD1 and significantly, the addition of two to four heptapeptides markedly increases the B1C8 disrupting activity: only approximately 30% of nuclei expressing F-CTD3 or F-CTD6 have intact B1C8 speckles.

The addition of 7, 20 or 26 heptapeptides to F-CTD6 does not further reduce the frequency of nuclei with intact B1C8 speckles, but the longer CTD segments (e.g. F-CTD13, F-CTD26 and F-CTD32) induce a more severe disruption of the B1C8 speckles than short CTD segments (not reflected by the histogram in Figure 2). Significantly, F-CTD52 induces a complete disruption of the B1C8 speckled pattern in nearly 100% of the transfected nuclei. A similar trend was observed with a nested set of CTD sequences linked to βGal. These data indicate that the speckled distribution of an SR splicing protein (B1C8) is incrementally disrupted by the stepwise addition of heptapeptide repeats to the fusion protein.

Example 5: Multiple SR splicing factors and Sm snRNPs redistribute from a speckled to a diffuse pattern in nuclei expressing CTD-derived proteins.

B1C8 is one of many SR family splicing proteins in speckle domains (reviewed by Fu, RNA 1:663-680 1995). To ascertain whether CTD-derived proteins alter the distribution of the SR proteins recognized by these reagents, the experiment described above was repeated except that mAb 3C5, mAb 104, mAb NM4 or NM22 was substituted for mAb B1C8. The results indicate that F-CTD52 disrupts the speckled staining pattern of all four antibodies.

Speckle domains are also enriched with other classes of splicing factors, such as Sm snRNPs and U-rich snRNAs. The preceding study showed that Pol IIo can be co-immunoprecipitated with antibodies directed at Sm snRNPs, so it was determined whether CTD-derived proteins induce Sm snRNP antigens to become dispersed. The Sm snRNPs were localized with mAb Y12 (an IgG), so the anti-Flag mAb M2

could not be used for double staining. This problem was addressed in two ways. In the first experiment, transfected cell nuclei were distinguished from untransfected nuclei by immunostaining with mAb H5 (IgM). This antibody recognizes phosphoepitopes on the CTD, and it stains nuclei expressing phosphorylated CTD-derived proteins much more intensely than control nuclei. In a second experiment, CV1 cells were transfected with pβGal-CTD52, and double stained with anti-βGal (rabbit IgG) and mAb Y12.

A nucleus expressing F-CTD52 is identified by the intense immunostaining with mAb H5. Three untransfected cell nuclei are identified by weaker mAb H5 immunostaining. A nucleus expressing βGal-CTD52 is identified by intense immunostaining with mAb anti-βGal, and three untransfected cell nuclei are identified by faint immunostaining with mAb anti-βGal. The distribution of F-CTD13 is revealed by red pseudocolor in three transfected cell nuclei. The distribution of p80 coilin is revealed by green pseudocolor in the three cells expressing F-CTD13, and in one untransfected cell. Sm antigens are observed in speckle domains of the untransfected nuclei, but they are diffusely distributed in the transfected cell nucleus. The βGal linked CTD-52 protein has a more striking effect on the Sm antigens. Immunostaining with anti-βGal reveals a brightly stained nucleus expressing the βGal-CTD52 protein, and three faintly stained control cell nuclei. Examination of the same cells stained with mAb Y12 reveals that the Sm antigens are distributed more diffusely in the transfected cell nucleus than in the untransfected cell nuclei.

Example 6: Expression of CTD-derived fusion proteins disrupts speckle domains, but not coiled bodies.

Coiled bodies (CBs) are dot-like nuclear domains that contain certain snRNPs and snRNAs that are also present in the speckle domains (Lamond and Carmo-Fonseca Trends in Cell Biology 3:198-204 (1993). Most cultured mammalian cells have 2-5 CBs, which are easily visualized by immunostaining with antibodies directed at the p80 coilin autoantigen (Andrade et al., 1993). Speckles and CBs both contain certain splicing components, but their composition is otherwise very different: Pol IIo and SR splicing factors are present in speckle domains, but they have not been reported in CBs. Similarly, CBs contain p80 coilin, fibrillarin and Nopp140, which have not been reported in speckle domains. Finally, transcriptional inhibitors and heat shock cause CBs to shrink and speckle domains to enlarge, suggesting distinct physiological roles for these two types of domains.

To ascertain whether the CTD-derived proteins disrupt the organization of CBs, each Flag-tagged CTD-derived protein was expressed transiently in CV1 cells, which were fixed and double immunostained with anti-p80 coilin and anti-Flag mAb M2. The results indicate that the distribution of p80-coilin is unaffected by CTD-derived proteins F-CTD52, F-CTD32, F-CTD26 and F-CTD13. In the example presented here, CBs are observed in a control cell nucleus as well as three nuclei expressing F-CTD13.

Example 7: Expression of F-CTD52 blocks the accumulation of spliced, but not unspliced, β-globin transcripts *in vivo*.

The transfection experiments described above, supports the hypothesis that Pol IIo associates with SR splicing factors and Sm snRNPs and via its CTD, but they do not provide evidence indicating a functional relationship between the CTD and pre-mRNA splicing. If the processes of transcription and splicing are linked via a CTD-mediated mechanism, then one might expect the CTD-derived proteins to interfere with transcription, splicing or both processes. CTD-derived proteins and β-globin transcripts were therefore co-expressed in the same nucleus. A series of double expression plasmids was created by inserting intact β-globin genes into pF-CTD52, pF-CTD13, pF-CTD6 and pF-CTD1 (Figure 3A). As controls, intact β-globin genes were inserted into pF-CTDless.1, pF-CTDless.3 and pβGal (Figure 3A). Each Flag-tagged CTD coding sequence (or control sequence) is under the control of a CMV promoter, whereas the β-globin gene is driven by its own promoter. The β-globin genes were also inserted in the opposite orientation relative to the Flag-tagged CTD coding sequences to control for possible *cis* effects (Figure 3B).

First, HeLa cells were transfected with pF-CTDless.1β-globin [+], pF-CTD13β-globin [+] or pF-CTD52β-globin [+]. Twenty four hours later, spliced and unspliced β-globin transcripts were quantitated by RT-PCR. PCR primers (P1 and P2) hybridize with sequences within exons 1 and 2, and therefore amplify a segment that includes intron 1 (Figure 3A). The PCR products corresponding to spliced and unspliced β-globin transcripts are 170 and 300 nucleotides, respectively. The results of

this experiment indicate that co-expression of F-CTD52 reduces the amount of spliced β-globin transcript compared to the control, F-CTDless.1. In contrast, slightly more unspliced β-globin transcript accumulates in cells co-expressing F-CTD52 than in the control cells. An intermediate effect is achieved by co-expressing F-CTD13.

This result was confirmed using an RNase protection assay. Here, a 343 nt protecting RNA probe was designed to hybridize with 203 nucleotides of the second β-globin exon and 73 nucleotides at the 3′ end of intron 1. Thus, unspliced β-globin transcripts protect 276 nucleotides, and spliced transcripts protect 203 nucleotides of the radiolabeled probe (Figure 7A). Similar amounts of spliced and unspliced β-globin transcripts are present in cells expressing the control protein; however, one observes no spliced β-globin RNA in cells co-expressing the FCTD52 protein. Significantly, this reduction is accompanied by an increase of unspliced β-globin transcript. Splicing is inhibited to a lesser degree by F-CTD13 than F-CTD52.

To control for possible *cis* effects between the β-globin gene and CMV-Flag-CTD transcription unit, their relative orientation on the plasmids were reversed. The resulting plasmid constructs (pF-CTDless.1β-globin [-], pF-CTD13β-globin [-] or pF-CTD52β-globin [-]) were transfected into HeLa cells, and the RNase protection assay was performed. Again, one observes a reduction of spliced β-globin RNA in cells co-expressing the F-CTD52 protein. A less severe inhibitory effect is achieved by co-expressing F-CTD13. More unspliced β-globin transcript accumulates in cells expressing CTD-derived proteins than in control cells. This result indicates that CTD-derived proteins do not

block in transcription by RNA polymerase II. Indeed, CTD-derived proteins selectively interfere with splicing.

The last series of experiments utilized a thalassemic β-globin gene that has a G to A transition at position 1 in intron 1. The thalassemic pre-mRNAs are spliced at three cryptic 5′ splice sites, each located upstream of the 343 nt RNA probe used in the RNase protection assay. All three cryptically spliced products should protect 203 nucleotides of the radiolabeled probe, because they all utilize the same 3′ splice site. The thalassemic gene was substituted for wild type i-CTDless.1β-globin [+], pF-CTD13β-globin [+] and pF-CTD52β-globin [+] (Figure 7C), the resulting plasmids were transfected into HeLa cells, and RNase protection experiments were performed as before. Splicing of this thalassemic transcript is particularly sensitive to the inhibitory effects of the CTD-derived proteins.

It was then determined whether the removal of heptapeptide repeats from F-CTD52 progressively decreases the inhibitory effect on *in vivo* splicing. To test this idea, HeLa cells were transfected with plasmids that co-express the β-globin^{thal} transcript and one of a nested set of CTD-derived proteins. An RNase protection assay was performed. The ratio of unspliced to spliced β-globin^{thal} transcripts is not significantly different in cells expressing F-CTD-less.3, βGal, F-CTD-less.1, or F-CTD1; however, this ratio increases progressively as one adds 6, 13 and 52 heptapeptide repeats to the fusion protein. Indeed, a comparison of spliced β-globin^{thal} transcripts in cells expressing F-CTD1, F-CTD6, F-CTD13 and F-CTD52 reveals a graded inhibition of splicing,

which correlates with the number of heptapeptides added to the fusion protein.

These studies show that CTD-derived proteins are phosphorylated *in vivo* and accumulate in the nucleus, where they disrupt splicing factor domains and interfere with pre-mRNA splicing. In agreement with these *in vivo* results, CTD heptapeptides were shown to specifically inhibit *in vitro* splicing reactions. Taken together, these studies provide evidence for a functional interaction between Pol II's CTD and the splicing process, and they strongly imply that transcription and pre-mRNA splicing are coordinated by a mechanism involving a phosphorylated form of the CTD.

We claim:

1. A peptide conjugate comprising:

(a) a peptide comprising at least two hexapeptide repeats having the formula: $YX_1PX_2X_3PX_4$, where Y is tyrosine, P is proline, and X can be any amino acid,

(b) a bioactive molecule, and

(c) a linker between the bioactive molecule covalently attached to the peptide.

2. The conjugate of claim 1 wherein the linker consists of one to two amino acids or a carbon chain of equivalent length.

3. The conjugate of claim 1 wherein the linker is attached at the N-terminus of the peptide, leaving a free carboxyl.

4. The conjugate of claim 1 wherein X_1, X_2 and X_3 are serine or threonine.

5. The conjugate of claim 1 wherein the bioactive molecule is selected from the group consisting of proteins or peptides, sugars, and nucleic acid sequences.

6. The conjugate of claim 5 wherein the nucleic acid sequences are selected from the group consisting of ribozymes, external guide sequences for RNAase P, antisense, aptamers, triplex forming oligonucleotides, nucleosides, nucleotides, genes, cDNA, mRNA, and RNA.

7. The conjugate of claim 1 comprising a heptapeptide selected from the group consisting of:

YSPTSPS YSPTSPN YTPTSPN YSPTSPA YTPQSPS
YEPRSPGG YSPTSPT YSPTSPK YTPTSPK YSPTTPK
YSPTSPV YSPTSPG YSLTSPA YTPSSPS YSPSSPS
YTPTSPS YSPSSPE YTPQSPT YSPSSPR.

8. The conjugate of claim 1 wherein the peptide is phosphorylated.

9. The conjugate of claim 1 further comprising a pharmaceutically acceptable carrier for administration to cells.

10. The conjugate of claim 1 further comprising a detectable label.

11. A method for delivering a bioactive compound to the nucleus of a cell comprising administering to the cell the conjugate of any of claims 1-10.

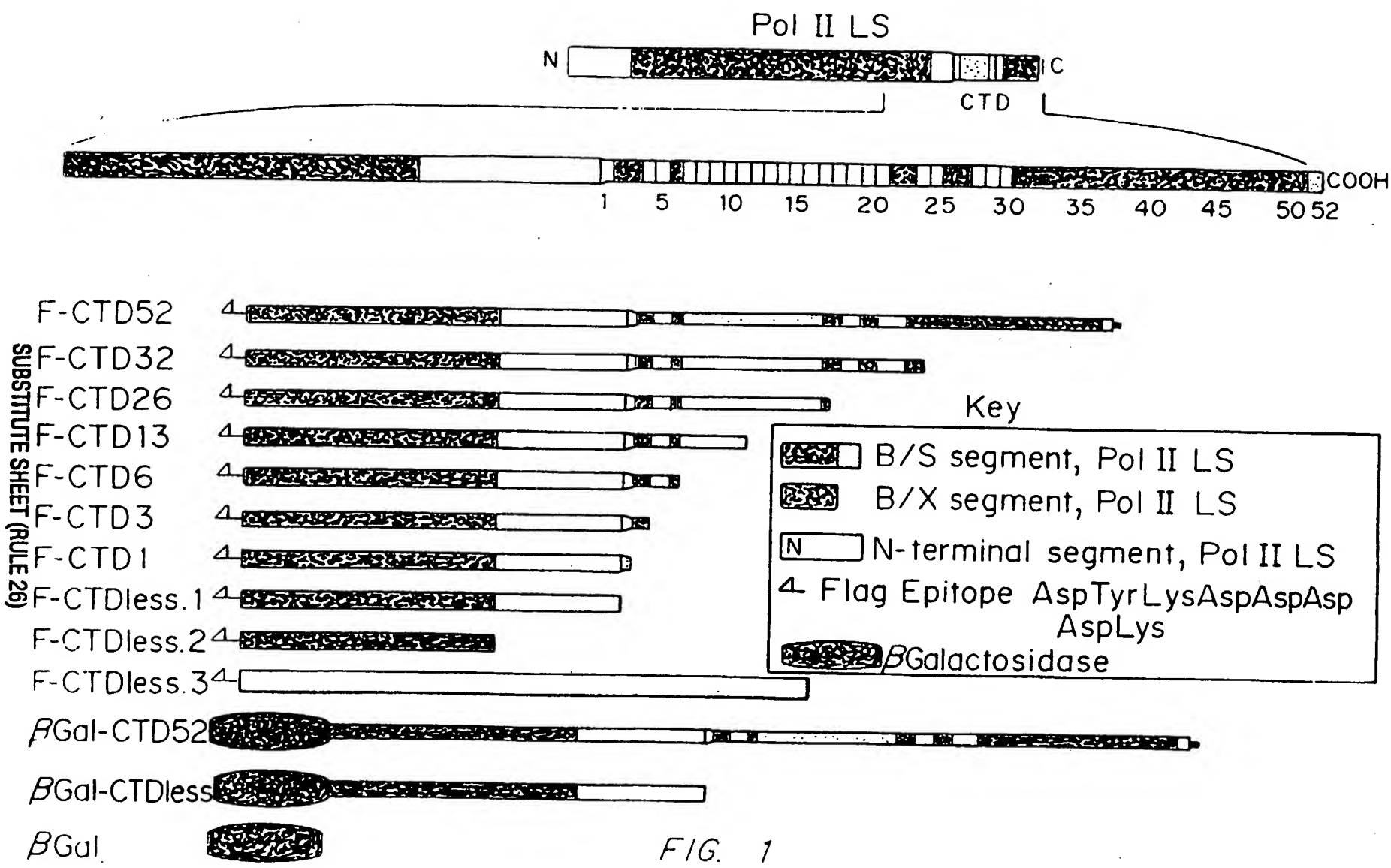


FIG. 1

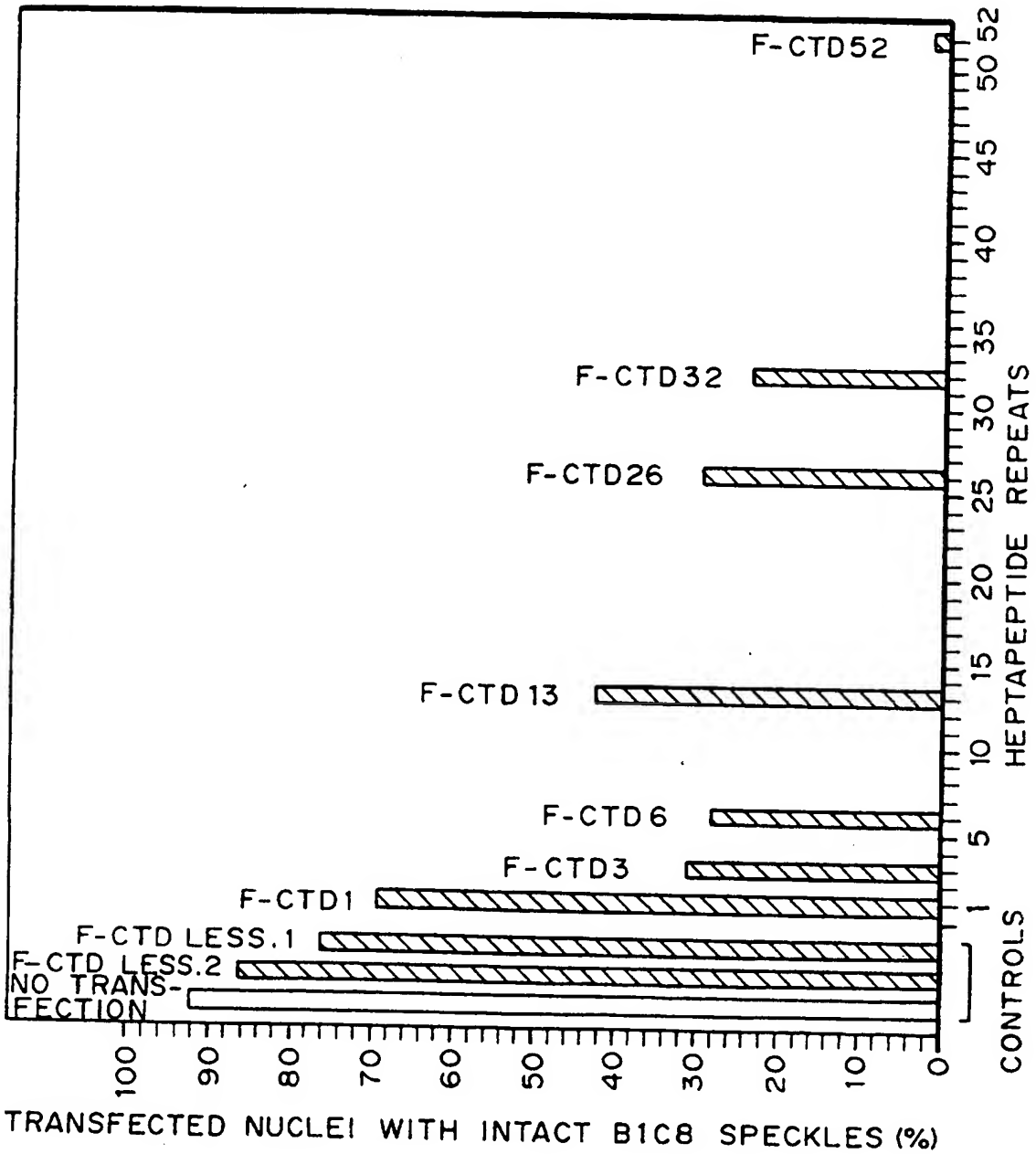

F-CTD52
F-CTD32
F-CTD26
F-CTD13
F-CTD6
F-CTD3
F-CTD1
F-CTD LESS.1
F-CTD LESS.2
NO TRANS-FECTION
52 50 45 40 35 30 25 20 15 10 5 1
HEPTAPEPTIDE REPEATS
CONTROLS
100 90 80 70 60 50 40 30 20 10 0
TRANSFECTED NUCLEI WITH INTACT B1C8 SPECKLES (%)

FIG. 2

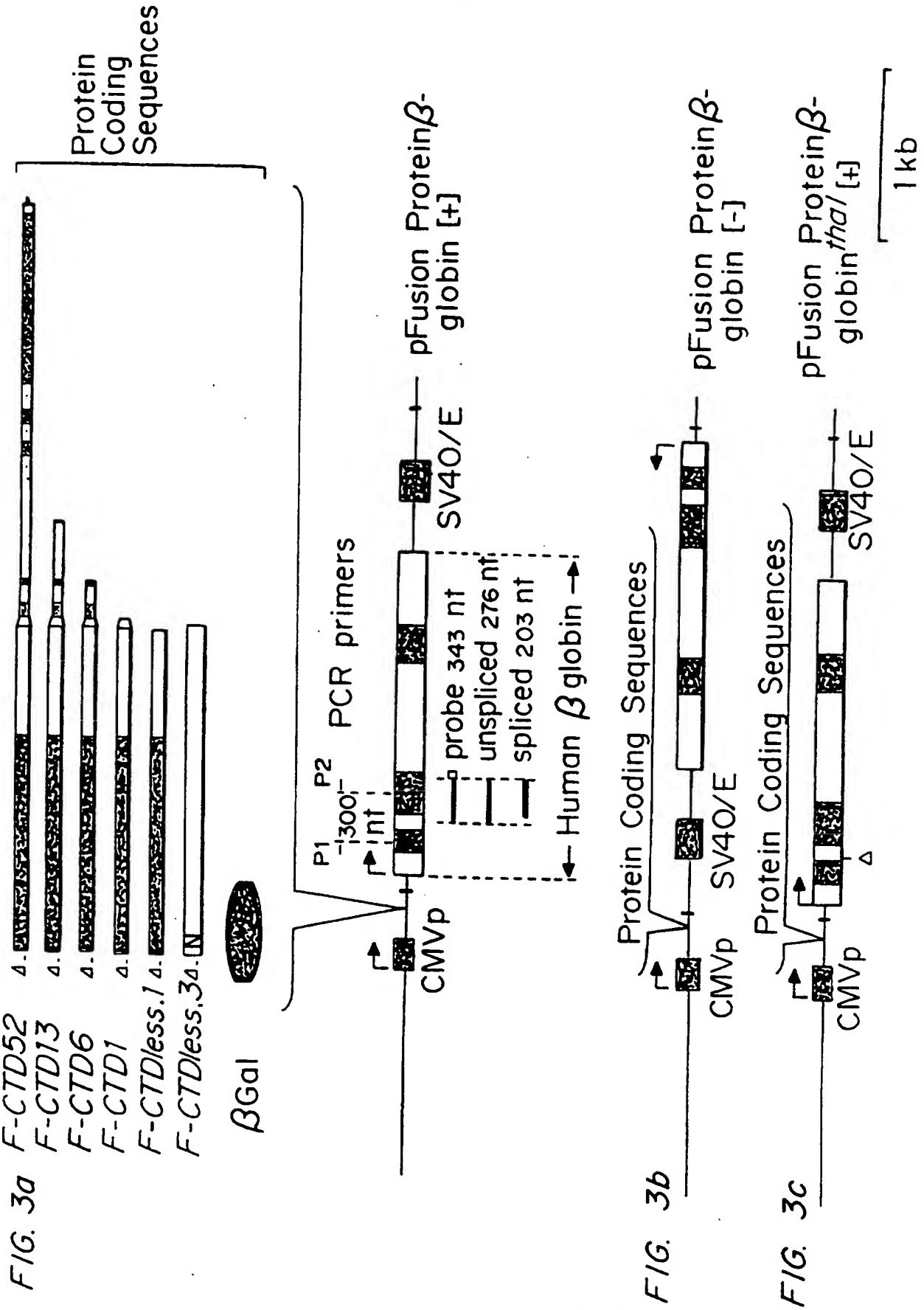


FIG. 3a

FIG. 3b

FIG. 3c